THE DEGREE OF CORRESPONDENCE BETWEEN TWO SERIES OF INDEX NUMBERS.

BY J. D. MAGEE, A.M., *Instructor in Economics, Western Reserve University.*

Our present day Economics is characterized by a great use of statistics. We are no longer satisfied with mere theoretical results if statistical verification is at all possible. One feature of the increased employment of statistics is the increasing use of index numbers. Sometimes they are used as the best means for displaying the relative changes of a series of definite quantities, as in the case of index numbers for exports or imports. But perhaps their most important use is to express relative changes in complicated data, which can be shown satisfactorily by no other method. This type is illustrated by the index numbers for wages computed by Bowley and by the various well known index numbers of general prices.

A new problem arises when, having index numbers for two sets of data, we attempt to get a quantitative expression for the closeness with which variations in one series are followed or accompanied by variations in the other series. Those who desire a quantitative expression for the relationship, and are not content with graphing the two sets of index numbers and speaking in general terms of the closeness of the relationship, have usually employed the Pearsonian Coefficient of Correlation. This device has been developed in the application of statistics to biological problems.

It is the purpose of this paper to question the applicability of this correlation coefficient to problems involving two series of index numbers and to suggest a possible method for dealing with the problem. Of course, the objections brought forward have no bearing on its applicability to biological problems for which it was originally designed.

To make the criticism clear it will be necessary to explain, in some detail, how the correlation coefficient is computed. We may take the form used by Yule.* Correlation coefficient $= r = \frac{\Sigma(xy)}{N\sigma_x\sigma_y}$ where x and y represent the deviations of

* Theory *of Statistics*, 174.

corresponding terms of the two series X and Y from their respective arithmetic means; where N is the number of terms in the series; and σ_x, σ_y are the standard deviations of the two series, that is, the square root of the quotient obtained by dividing the sum of the squares of the deviations of the various items of one series from the arithmetic mean of that series by the number of items (deviation = the item − the mean). The arithmetic mean is, of course, merely the sum of the items of the series divided by the number of items. If the correlation is perfect, the coefficient has the value $+1$; if there is perfect negative correlation the coefficient has the value -1. It is seen that the correlation coefficient is concerned with the admodality of the two series, that is, the manner in which the terms are grouped about the means of the series.

One objection may now be stated. We have seen that index numbers give us relative changes and that when we compare two series we are interested in seeing how the changes in one series correspond to changes in the other, but the correlation coefficient takes no account of the order in which the items of the series occur. The two series would have the same correlation

100	80	and	100	80
75	95		130	55
130	55		75	95
90	120		90	120

coefficient, though in the first case an increase or decrease of the one set is always accompanied by the reverse in the other, and in the second case in one instance there is an increase in both sets of numbers. There are six possible orders in which the above four pairs of numbers may be arranged (irrespective of the pair which is taken to start with). Of these six, two orders give for every increase or decrease of one series the reverse in the other. In four orders there is one case each in which an increase or decrease in one series is accompanied by an increase or decrease in the other. The formula for the correlation coefficient considers the relation of the various pairs of values in the series to their respective means but not at all their relation to the pairs of values which precede or follow them. Since the relation of the pairs of numbers to those nearest them is the significant thing in dealing with

index numbers and since the correlation coefficient is not affected by changes in this relationship, the correlation coefficient does not appear to be a good means for showing the degree of relationship between two sets of index numbers. In many cases the series of index numbers refer to various periods of time. The above objection is equivalent to saying that the correlation coefficient disregards the time element altogether.*

In order to make clear the next objection it will be necessary to examine some of the curious algebraic properties of the correlation coefficient. The first property is that the correlation coefficient of two series of index numbers is not changed if we add a constant to each term of one series. The reason is simple. If we add a constant k to each term of the series, the mean of the series will be increased by k. The standard deviation will not be changed. Each term and also the mean has been increased by k, so the difference between them will remain the same. By exactly the same reasoning the xy will not be changed, since x represents the difference between the term and the mean. Of course, we have not changed the number of terms, N. So there is no change whatever in the formula when we add a constant to each term of one series. Analogous reasoning would show that we may substract a constant from each of the terms of one series without changing the value of the correlation coefficient. It is also obvious that we may add or substract a constant (not necessarily the same one) to or from both series of numbers without changing the value of the correlation coefficient.

The second algebraic property of the correlation coefficient to be noted is that we may multiply each term of one of the series by the same number without changing the value of the correlation coefficient. Suppose we multiply each term by l; then the mean will be l times the former mean. Both the mean and the terms will be l times what they were formerly and so their differences will be l times the former difference. When the differences are squared, we will have l^2 times the former differences squared. Taking the sum of the squares of the deviations of the various terms and divid-

* Cf., Fisher, *Purchasing Power of Money*, p. 279.

ing by the number of terms and then extracting the square-root, our new standard deviation will be l times the former one. The x's in the numerator of the formula of the correlation coefficient, which represent the deviations from the mean will now all be multiplied by l since both the mean and the terms are l times their former value. Each being l times the former value, the sum will be l times the former sum. The result, then, of multiplying each term of the series by l is to multiply both numerator and denominator of the fraction which is the correlation coefficient by l. The l's may be cancelled and so the value of the correlation coefficient is not changed. It is obvious that we may multiply each of the two series by a constant multiplier (not necessarily the same) without changing the value of the correlation coefficient. Having examined these algebraic properties of the correlation coefficient, we shall now see how they affect its applicability to the problem of the relationship of two series of index numbers.

The possibility of multiplying either or both series by some multiplier without changing the value of the correlation coefficient, is evidently a point in its favor. It will be remembered that our index numbers are merely relative. The ratio of the quantity in one year to that of the next year is given by the ratio of the index number of the first year to the index number of the second year. Suppose the index number is for Imports into the United States, we may express it as follows: $\frac{\text{Imports 1855}}{\text{Imports 1860}} = \frac{100}{137.2}$. It is evident that multiplying each of the index numbers, e. g., by 2, will leave the value of the ratio the same. If then the new series expresses the same relations as before, the correlation coefficient with some other series should remain the same.

The fact, however, that a given quantity may be added to or substracted from the terms of our series without changing the value of the correlation coefficient, is perhaps the strongest argument against using it for testing the relationship between two sets of index numbers; for when the same number is added to each of a series of index numbers, they no longer express the same relations. This proposition may be made clear by a

simple example. If there is a series of index numbers 50, 60, 80, the ratio of the first to the second term is 5 to 6 and of the second to the third term, 3 to 4. Now if 100 be added to each term the series becomes, 150, 160, 180. Here the ratio of the first to the second term is 15 to 16 and of the second to the third term, 8 to 9. It is seen that the ratios between the terms are entirely changed. The objection then is plain. The correlation coefficient would show that these two series were perfectly correlated, while as a matter of fact they do not show the same relative changes. Various combinations of adding a constant or multiplying by a constant might be made. The correlation as found by the correlation coefficient is perfect for the following sets of numbers:

110	150
160	225
130	180
120	165

The relation may not be evident at first, but it will be discovered that in each case the term of the second series is $\frac{3}{2}$ of ten less than the corresponding term of the first series. The situation, then, is that the correlation coefficient will indicate perfect correlation between two sets of index numbers if they show the same relative changes, but when the correlation coefficient indicates perfect correlation, the two series may not show the same relative changes at all. Surely this uncertainty is an indication of the unfitness of the correlation coefficient to test the relationship between two sets of index numbers. A chemical test would not be considered of great value which always indicated the presence of an element but also gave the same reaction in other cases when the element was not present.

Our objections to the Pearsonian Correlation Coefficient as a means of testing the relationship between two series of index numbers are, to sum up: (1) it entirely disregards the element of time which in most problems in which index numbers are used, is of prime importance; and (2) the result obtained is not definite, if the relative changes are the same we get perfect correlation, but perfect correlation does not always mean that the relative changes are the same. It may be well to repeat that these objections apply only to the use of the

correlation coefficient in connection with index numbers and not at all to the general use and especially not to its use in Biology.

Having brought forward these objections to the use of the correlation coefficient as a means of showing the relationship between two series of index numbers, it may be well to suggest a better means. We may indicate the way a figure can be found for the closeness of the relationship between two series of index numbers. This figure may be called the Degree of Correspondence to avoid the confusion of using the term correlation coefficient for what is obtained in another way. We assume that we have two series of index numbers which represent the relative changes in two sets of data, usually for a series of years. We are interested in finding out whether an increase or decrease in one is associated with a like change in the other. If this be the case we say there is positive correspondence. If they both increase or decrease in the same ratio, there is perfect positive correspondence. If one increases and the other decreases there is negative correspondence. If the one increases in the same ratio as the other decreases, there is perfect negative correspondence. If one changes and the other remains constant, there is no, or zero, correspondence.

With these things understood, we may figure the degree of correspondence in two ways. First, roughly, we may examine our series and for every case of positive correspondence we put down a +1 and for every case of negative correspondence, a –1 and for every case where one is constant while the other changes, a 0. Then the arithmetic mean of these numbers may be taken as the rough value of the degree of correspondence between the two series. It is evident that the figure may vary from +1, which would indicate perfect positive correspondence (considering only the direction of the change, not the amount) to –1, which would be perfect negative correspondence (of direction of change). The case where there is an equal number of +1's and –1's and so the degree of correspondence would be 0, is of interest. In many arguments where the method here suggested has been practically used, however, (the number of agreements and disagreements is

usually pointed out) it has been held that some relationship is shown if half of the cases are agreements. Yet, if there were but two options and the series were totally unrelated, mere chance would give approximately half agreements and half disagreements, the nearness of the approximation depending upon the number of cases considered. It is evident that in talking of correspondence we wish to imply something more than what might be the result of chance.

Second, we may get a more accurate form of the degree of correspondence. To obtain this it is necessary to substitute for the $+1$ or -1 as used in the rough method a more exact expression. The following procedure may be adopted. We are dealing, of course, with pairs of index numbers. By multiplying one or both pairs by factors we do not change the relations expressed but we may make the first numbers of the two pairs agree. Next we may get the amount of change in each case and take as the degree of correspondence the fraction which has the smaller change for the numerator and the larger change for the denominator. An example will serve to make the process plain. We start with a part of a table of index numbers which is as follows: $\begin{matrix}60 & 50.\\ 72 & 55\end{matrix}$ The first step is to make the first two numbers agree. We may do this either by multiplying each term of the second set by $\frac{6}{5}$ thus giving $\begin{matrix}60 & 60\\ 72 & 66\end{matrix}$ or by multiplying each term of the first set by $\frac{5}{6}$ thus getting $\begin{matrix}50 & 50.\\ 60 & 55\end{matrix}$ The degree of correspondence would then be $+\frac{1}{2}$ ($\frac{6}{12}$ in the first case, $\frac{5}{10}$ in the second case). The sign is plus because the changes are in the same direction. If we had $\begin{matrix}50 & 50,\\ 60 & 35\end{matrix}$ the degree of correspondence would be $-\frac{10}{15}$ or, reducing, $-\frac{2}{3}$. The sign is minus because the changes are in opposite directions. By treating each successive change in this manner and then taking the arithmetic mean of the results we obtain what may be called the Degree of Correspondence in the more accurate form. In brief, the suggested Degree of Correspondence is the average correspondence

(computed in the manner suggested above) shown by successive terms of two series of index numbers.

One example of the application of the proposed Degree of Correspondence may be given. Irving Fisher has attempted a statistical proof of the quantity theory of money. He gives* two series of index numbers for the years 1896 to 1909. Both series represent price changes; one is obtained directly from price statistics and the other is obtained indirectly by solving the equation of exchange for the price level. Fisher does not consider the Pearsonian Correlation Coefficient satisfactory, but he finds the value of it for his figures because it has been used to test previous attempts at statistical verification of the quantity theory of money. Applying it in two different ways he gets correlations of +.97 and +.57.† He prefers the latter figure. Below is given the application of the proposed method.

Year.	Index Numbers of Prices.		Degree of Correspondence.		
	Direct.	Indirect.	Positive.	Zero.	Negative.
1896	63	54	...	..	.43
1897	64	52	...	..	...
1898	66	56	.41	..	...
1899	74	69	.52	..	...
1900	80	68	...	..	.18
1901	84	76	.43	..	...
1902	89	82	.75	..	...
1903	87	75	.26	..	...
1904	85	81	...	..	.29
1905	91	83	.35	..	...
1906	97	90	.78	..	...
1907	97	86	...	0	...
1908	92	87	...	..	.23
1909	100	100	.58	..	...
			+4.08		−1.13 = 2.95 ÷ 13 = +.23

The degree of correspondence is then +.23. The degree of correspondence between the series obtained by taking alternate years is much higher. If we start with the first year, it is +.67; if we start with the second year, it is +.34. If we take merely the first and last terms of the series, the degree of correspondence is +.69.

* *Purchasing Power of Money* p. 293.

† *Ibid*, pp. 294–5.